IN THE UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF TEXAS
TYLER DIVISION

SPIDER SEARCH ANALYTICS LLC	§	
	§	
Plaintiff,	§	**CIVIL ACTION NO.**
	§	
v.	§	**JURY TRIAL DEMANDED**
	§	
CAREERBUILDER, LLC,	§	
	§	
Defendant.	§	

COMPLAINT FOR INFRINGEMENT OF PATENT

COMES NOW, Plaintiff Spider Search Analytics LLC ("SSA" or Plaintiff), through the undersigned attorneys, and respectfully alleges, states, and prays as follows:

NATURE OF THE ACTION

1. This is an action for patent infringement under the Patent Laws of the United States, Title 35 United States Code ("U.S.C.") to prevent and enjoin defendant CareerBuilder, LLC, (hereinafter "Defendant") from infringing and profiting, in an illegal and unauthorized manner and without authorization and/or of the consent from SSA, from U.S. Patent No. 7,454,430 (the "'430 patent", attached hereto as Exhibit "A") (the "Patent-in-Suit") pursuant to 35 U.S.C. § 271, and to recover damages, attorney's fees, and costs.

THE PARTIES

2. Plaintiff SSA is a Texas entity with its principal place of business at 101 E. Park Blvd, Suite 600, Plano, Texas 75074.

3. Upon information and belief, Defendant is a company organized and existing under the laws of the State of Delaware, having a principal place of business at 200 N. LaSalle

St., Suite 1100, Chicago, IL 60601. Upon information and belief, Defendant may be served with process at CT Corporation System, 1999 Bryan St., Ste. 900, Dallas, TX 75201.

JURISDICTION AND VENUE

4. The Court has subject matter jurisdiction over this action pursuant to 28 U.S.C. §§ 1331 and 1338(a) because the action arises under the Patent Laws of the United States, 35 U.S.C. §§ 1 *et seq*.

5. This Court has personal jurisdiction over Defendant by virtue of its systematic and continuous contacts with this jurisdiction, including having the right to transact business in Texas, as well as because of the injury to SSA, and the cause of action SSA has risen, as alleged herein.

6. Defendant is subject to this Court's personal jurisdiction pursuant to due process and/or the Texas Long Arm Statute, due at least to its substantial business and purposeful availment of this forum, including: (i) at least a portion of the infringements alleged herein; and (ii) regularly doing or soliciting business, engaging in other persistent courses of conduct, and/or deriving substantial revenue from goods and services provided to individuals in Texas and in this judicial district.

7. Upon information and belief, Defendant, directly and/or through its employees or agents, and/or its customers, uses products, as defined below, that perform each and every step of at least one claim of the '430 patent with the knowledge and/or understanding that such products are used or will be used in this District. Upon information and belief, Defendant has engaged in substantial and not isolated activity within this District. Therefore, exercise of jurisdiction over Defendant will not offend traditional notions of fair play and substantial justice. Such an exercise is consistent with the Texas long-arm statute.

8. Venue is proper in this judicial district pursuant to 28 U.S.C. §§ 1391 and 1400(b) because Defendant is subject to personal jurisdiction in this district, has regularly conducted business in this judicial district and certain of the acts complained of herein occurred in this judicial district.

FACTUAL ALLEGATIONS

9. On November 18, 2008, the United States Patent and Trademark Office ("USPTO") duly and legally issued the '430 patent, entitled "System and method for facts extraction and domain knowledge repository creation from unstructured and semi-structured documents" after a full and fair examination. (Exhibit A).

10. SSA is presently the owner of the patent, having received all right, title and interest in and to the '430 patent from the previous assignee of record. SSA possesses all rights of recovery under the '430 patent, including the exclusive right to recover for past infringement.

11. The '430 patent contains three (3) independent claims and twenty-four (24) dependent claims.

12. The '430 patent claims, *inter alia*, a method for building a deep web crawler.

13. Defendant uses, *inter alia*, a method that performs each and every step of at least one claim of the '430 patent.

DEFENDANT'S PRODUCTS

14. In accordance with claim 10 of the '430 patent Defendant uses a method for building a deep web crawler. For example, Defendant is a company that uses the web crawler "Scrapy" (the "Accused Product") in order to "scrape job offers from many sites".[1] The Accused Product allows customers to build web crawlers that are capable of crawling any website

[1] https://scrapy.org/companies/, last visited May 8, 2017.

(including, upon information and belief, websites from the "deep web" and that use AJAX for dynamic pages).

careerbuilder®

CareerBuilder: Uses Scrapy to scrape job offers from many sites.

Source: https://scrapy.org/companies/

15. In accordance with claim 10 of the '430 patent Defendant utilizes scout crawling rules to collect dynamic pages.[2]

Scrapy Documentation, Release 1.3.0

site. Rules objects are described below. If multiple rules match the same link, the first one will be used, according to the order they're defined in this attribute.

This spider also exposes an overrideable method:

parse_start_url (*response*)
This method is called for the start_urls responses. It allows to parse the initial responses and must return either an `Item` object, a `Request` object, or an iterable containing any of them.

Crawling rules

class `scrapy.spiders.`**Rule** (*link_extractor*, *callback=None*, *cb_kwargs=None*, *follow=None*, *process_links=None*, *process_request=None*)

`link_extractor` is a *Link Extractor* object which defines how links will be extracted from each crawled page.

`callback` is a callable or a string (in which case a method from the spider object with that name will be used) to be called for each link extracted with the specified link_extractor. This callback receives a response as its first argument and must return a list containing `Item` and/or `Request` objects (or any subclass of them).

> **Warning:** When writing crawl spider rules, avoid using `parse` as callback, since the `CrawlSpider` uses the `parse` method itself to implement its logic. So if you override the `parse` method, the crawl spider will no longer work.

`cb_kwargs` is a dict containing the keyword arguments to be passed to the callback function.

`follow` is a boolean which specifies if links should be followed from each response extracted with this rule. If `callback` is None `follow` defaults to `True`, otherwise it defaults to `False`.

`process_links` is a callable, or a string (in which case a method from the spider object with that name will be used) which will be called for each list of links extracted from each response using the specified `link_extractor`. This is mainly used for filtering purposes.

`process_request` is a callable, or a string (in which case a method from the spider object with that name will be used) which will be called with every request extracted by this rule, and must return a request or None (to filter out the request).

---

[2] *See* Scrapy Documentation, Release 1.3.0, p. 34.
Source: https://media.readthedocs.org/pdf/scrapyzero/1.3.0/scrapyzero.pdf.

16. In accordance with claim 10 of the '430 patent Defendant utilizes an analyzer and extractor to determine underlying structure of queries. For example, Defendant provides a list of Start URLs which are added into the crawl queue. Scrapy's analyzer then utilizes a CSS selector to analyze the HTML elements on crawled pages to determine the underlying structure of queries including but not limited to the links on those pages. The analyzer uses extractor code to extract the HTML and identify hyperlinks for this purpose. The qualifying links are then fed into the crawl queue.[3]

17. In accordance with claim 10 of the '430 patent Defendant generates instructions for a harvester, wherein the harvester provides requests to a server and collects available pages from the server. For example, once the analyzer adds all the Start URLs and the relevant links that were found on the pages corresponding to the Start URLs, to the crawl queue, Defendant generates the instructions for a harvester (i.e., a spider) and the harvester begins transmitting requests to the server for the web pages represented by the entries in rule set.[4]

18. The elements described in paragraphs 14-17 are covered by at least claim 10 of the '430 patent.

INFRINGEMENT OF THE '430 PATENT

19. Plaintiff realleges and incorporates by reference the allegations set forth in paragraphs 1 to 18.

20. In violation of 35 U.S.C. § 271, Defendant is now, and has been directly infringing the '430 patent.

21. Defendant has had knowledge of infringement of the '430 patent at least as of the service of the present complaint.

---

[3] *See* Scrapy Documentation, Release 1.3.0, pg. 6.
[4] *See* https://doc.scrapy.org/en/1.3/topics/spiders.html.

22. Defendant has directly infringed and continues to directly infringe at least claim 10 of the '430 patent by using the Accused Product without authority in the United States, and will continue to do so unless enjoined by this Court. For example, Defendant's website, which is available in Texas, utilizes and implements the Accused Product and performs each and every step recited in claim 10 of the '430 patent.[5] As a direct and proximate result of Defendant's direct infringement of the '430 patent, Plaintiff has been and continues to be damaged.

23. By engaging in the conduct described herein, Defendant has injured SSA and is thus liable for infringement of the '430 patent, pursuant to 35 U.S.C. § 271.

24. Defendant has committed these acts of infringement without license or authorization.

25. As a result of Defendant's infringement of the '430 patent, SSA has suffered monetary damages and is entitled to a monetary judgment in an amount adequate to compensate for Defendant's past infringement, together with interests and costs.

26. SSA will continue to suffer damages in the future unless Defendant's infringing activities are enjoined by this Court. As such, SSA is entitled to compensation for any continuing and/or future infringement up until the date that Defendant is finally and permanently enjoined from further infringement.

DEMAND FOR JURY TRIAL

27. SSA demands a trial by jury of any and all causes of action.

PRAYER FOR RELIEF

WHEREFORE, SSA prays for the following relief:

1. That Defendant be adjudged to have infringed the Patent-in-Suit directly, literally and/or under the doctrine of equivalents;

---

[5] https://scrapy.org/companies/.

2. That Defendant, its officers, directors, agents, servants, employees, attorneys, affiliates, divisions, branches, parents, and those persons in active concert or participation with any of them, be permanently restrained and enjoined from directly infringing the Patent-in-Suit;

3. An award of damages pursuant to 35 U.S.C. §284 sufficient to compensate SSA for the Defendant's past infringement and any continuing or future infringement up until the date that Defendant is finally and permanently enjoined from further infringement, including compensatory damages;

4. An assessment of pre-judgment and post-judgment interest and costs against Defendant, together with an award of such interest and costs, in accordance with 35 U.S.C. §284;

5. That Defendant be directed to pay enhanced damages, including SSA's attorneys' fees incurred in connection with this lawsuit pursuant to 35 U.S.C. §285; and

6. That SSA have such other and further relief as this Court may deem just and proper.

Dated: May 16, 2017

Respectfully Submitted,

By: */s/Eugenio J. Torres-Oyola*
Eugenio J. Torres-Oyola
USDC No. 215505
Ferraiuoli LLC
221 Plaza, 5th Floor
221 Ponce de León Avenue
San Juan, PR 00917
Telephone: (787) 766-7000
Facsimile: (787) 766-7001
Email: etorres@ferraiuoli.com

Jean G. Vidal Font
USDC No. 227811
Ferraiuoli LLC
221 Plaza, 5th Floor
221 Ponce de León Avenue
San Juan, PR 00917
Telephone: (787) 766-7000
Facsimile: (787) 766-7001
Email: jvidal@ferraiuoli.com

ATTORNEYS FOR PLAINTIFF
SPIDER SEARCH ANALYTICS LLC